draft – 2011-09-01

Analyzing Inconsistencies in Probabilistic Conditional Knowledge Bases using Continuous Inconsistency Measures

Matthias Thimm

Technische Universität Dortmund, Germany

Abstract. Probabilistic conditional logic is a knowledge representation formalism that uses probabilistic conditionals (if-then rules) to model uncertain and incomplete information. By applying the principle of maximum entropy one can reason with a set of probabilistic conditionals in an information-theoretical optimal way, provided that the set is consistent. As in other fields of knowledge representation, consistency of probabilistic conditional knowledge bases is hard to ensure if their size increases or multiple sources contribute pieces of information. In this paper, we discuss the problem of analyzing and measuring inconsistencies in probabilistic conditional logic by investigating inconsistency measures that support the knowledge engineer in maintaining a consistent knowledge base. An inconsistency measure assigns a numerical value to the severity of an inconsistency and can be used for restoring consistency. Previous works on measuring inconsistency consider only qualitative logics and are not apt for quantitative logics because they assess severity of inconsistency without considering the probabilities of conditionals. Here, we investigate *continuous* inconsistency measures which allow for a more fine-grained and continuous measurement.

1 Introduction

Inconsistencies arise easily when experts share their beliefs in order to build a joint knowledge base. Although these inconsistencies often affect only a little portion of the knowledge base or emerge from only little differences in the experts' beliefs, they cause severe damage. In particular, for knowledge bases that use classical logic for knowledge representation, inconsistencies render the whole knowledge base useless, due to the well-known principle *ex falso quodlibet*. Therefore reasoning under inconsistency is an important field in knowledge representation and reasoning and there are basically two paradigms for approaching this issue. On the one hand one can live with inconsistencies and develop reasoning mechanisms that allow for consistent inference in the presence of inconsistent information, cf. e. g. *paraconsistent* and *default logics* [8]. On the other hand one can rely on classical inference mechanisms and ensure that knowledge bases are consistent, cf. e. g. approaches to *belief revision* and *information fusion* [2]. In this paper we employ probabilistic conditional logic [5] for knowledge representation. The basic notion of probabilistic conditional logic is that of a *probabilistic* *conditional* which has the form $(\psi \mid \phi)[d]$ with the commonsense meaning "if ϕ is true then ψ is true with probability d". A popular choice for reasoning with sets of probabilistic conditionals is model-based inductive reasoning based on the principle of maximum entropy [6, 5]. However, a prerequisite for applying this principle is the consistency of the set, i. e. the existence of at least one probability function that satisfies all probabilistic conditionals.

In this paper we investigate the issue of inconsistency in probabilistic conditional logic from an analytical perspective. One way to analyze inconsistencies is by measuring them. For the framework of classical logic, several approaches to analyze and measure inconsistency have been proposed—see e. g. [3]—and it is straightforward to apply those measures to the framework of probabilistic conditional logic [11]. However, those approaches do not grasp the nuances of probabilistic knowledge and allow only for a very coarse assessment of the severity of inconsistencies. In particular, those approaches do not take the crucial role of probabilities into account and exhibit a discontinuous behavior in measuring inconsistency. That is, a slight modification of the probability of a conditional in a knowledge base may yield a discontinuous change in the value of the inconsistency. In this paper, we consider measuring inconsistency in probabilistic conditional logic and continue previous work [10] in several aspects. First, we propose several novel principles for inconsistency measurement. Second, we pick up an extended logical formalization [7] of the inconsistency measure proposed in [10] and define a family of inconsistency measures based on minimizing the p-norm distance of a knowledge base to consistency. Third, we propose a novel compound measure that solves an issue with the previous measure and investigate its properties.

The rest of this paper is organized as follows. In Sec. 2 we give a brief overview on probabilistic conditional logic and continue in Sec. 3 with presenting a set of rationality postulates for continuous inconsistency measurement. We propose a family of inconsistency measures and a compound measure in Sec. 4 and analyze their properties in Sec. 5. We briefly review related work in Sec. 6 and conclude with a summary in Sec. 7.

2 Probabilistic Conditional Logic

Let At be a propositional signature, i. e. a finite set of propositional atoms. Let $\mathcal{L}(\mathsf{At})$ be the corresponding propositional language generated by the atoms in At and the connectives $\wedge$ (*and*), $\vee$ (*or*), and $\neg$ (*negation*). For $\phi, \psi \in \mathcal{L}(\mathsf{At})$ we abbreviate $\phi \wedge \psi$ by $\phi\psi$ and $\neg\phi$ by $\overline{\phi}$. The symbols $\top$ and $\bot$ denote *tautology* and *contradiction*, respectively. We use *possible worlds*, i. e. syntactical representations of *truth assignments*, for interpreting sentences in $\mathcal{L}(\mathsf{At})$. A possible world ω is a complete conjunction, i. e. a conjunction that contains for each $a \in \mathsf{At}$ either a or $\neg a$. Let $\Omega(\mathsf{At})$ denote the set of all possible worlds. A possible world $\omega \in \Omega(\mathsf{At})$ *satisfies* an atom $a \in \mathsf{At}$, denoted by $\omega \models a$ if and only if a positively appears in ω. The entailment relation $\models$ is extended to arbitrary formulas in

$\mathcal{L}(\mathsf{At})$ in the usual way. Formulas $\psi, \phi \in \mathcal{L}(\mathsf{At})$ are *equivalent*, denoted by $\phi \equiv \psi$, if and only if $\omega \models \phi$ whenever $\omega \models \psi$ for every $\omega \in \Omega(\mathsf{At})$.

The central notion of probabilistic conditional logic [5] is that of a *probabilistic conditional*.

Definition 1. *If $\phi, \psi \in \mathcal{L}(\mathsf{At})$ with $d \in [0,1]$ then $(\psi \,|\, \phi)[d]$ is called a* probabilistic conditional.

A probabilistic conditional $c = (\psi \,|\, \phi)[d]$ is meant to describe a probabilistic *if-then* rule, i. e., the informal interpretation of c is that "If ϕ is true then ψ is true with probability d" (see below). If $\phi \equiv \top$ we abbreviate $(\psi \,|\, \phi)[d]$ by $(\psi)[d]$. Further, for $c = (\psi \,|\, \phi)[d]$ we denote with $\mathsf{head}(c) = \psi$ the *head* of c, with $\mathsf{body}(c) = \phi$ the *body* of c, and with $\mathsf{prob}(c) = d$ the *probability* of c. Let $\mathcal{C}(\mathcal{L}(\mathsf{At}))$ denote the set of all probabilistic conditionals with respect to $\mathcal{L}(\mathsf{At})$.

Definition 2. *A* knowledge base $\mathcal{K}$ *is an ordered finite multi-subset of $\mathcal{C}(\mathcal{L}(\mathsf{At}))$, i. e. it holds that $\mathcal{K} = \langle c_1, \ldots, c_n \rangle$ for some $c_1, \ldots, c_n \in \mathcal{C}(\mathcal{L}(\mathsf{At}))$.*

We impose an ordering on the conditionals in a knowledge base $\mathcal{K}$ only for technical convenience. The order can be arbitrary and has no further meaning other than to enumerate the conditionals of a knowledge base in an unambiguous way. For similar reasons we allow a knowledge base to contain the same probabilistic conditional more than once. We come back to the reasons for these design choices later. For knowledge bases $\mathcal{K} = \langle c_1, \ldots, c_n \rangle$, $\mathcal{K}' = \langle c'_1, \ldots, c'_m \rangle$ and a probabilistic conditional c we define $c \in \mathcal{K}$ via $c \in \{c_1, \ldots, c_n\}$, $\mathcal{K} \subseteq \mathcal{K}'$ via $\{c_1, \ldots, c_n\} \subseteq \{c'_1, \ldots, c'_m\}$, and $\mathcal{K} = \mathcal{K}'$ via $\{c_1, \ldots, c_n\} = \{c'_1, \ldots, c'_m\}$. The union of belief bases is defined via concatenation.

Semantics are given to probabilistic conditionals by *probability functions* on $\Omega(\mathsf{At})$. Let $\mathcal{F}$ denote the set of all probability functions $P : \Omega(\mathsf{At}) \to [0,1]$. A probability function $P \in \mathcal{F}$ is extended to formulas $\phi \in \mathcal{L}(\mathsf{At})$ via

$$P(\phi) = \sum_{\omega \in \Omega(\mathsf{At}), \omega \models \phi} P(\omega) \quad .$$

If $P \in \mathcal{F}$ then P satisfies a probabilistic conditional $(\psi \,|\, \phi)[d]$, denoted by $P \models^{pr} (\psi \,|\, \phi)[d]$, if and only if $P(\psi\phi) = dP(\phi)$. Note that we do not define probabilistic satisfaction via $P(\psi \,|\, \phi) = {}^{P(\psi\phi)}/_{P(\phi)} = d$ in order to avoid a case differentiation for $P(\phi) = 0$, cf. [6]. A probability function P satisfies a knowledge base $\mathcal{K}$ (or is a model of $\mathcal{K}$), denoted by $P \models^{pr} \mathcal{K}$, if and only if $P \models^{pr} c$ for every $c \in \mathcal{K}$. Let $\mathsf{Mod}(\mathcal{K})$ be the set of models of $\mathcal{K}$. If $\mathsf{Mod}(\mathcal{K}) = \emptyset$ then $\mathcal{K}$ is *inconsistent*.

Example 1. Consider the knowledge base

$$\mathcal{K} = \langle (f \,|\, b)[0.9], (b \,|\, p)[1], (f \,|\, p)[0.1] \rangle$$

with the intuitive meaning that birds (b) usually (with probability 0.9) fly (f), that penguins (p) are always birds, and that penguins usually do not fly (only with probability 0.1). The knowledge base $\mathcal{K}$ is consistent as for e. g. $P \in \mathcal{F}$ with

$$
\begin{array}{llll}
P(bfp) = 0.005 & P(bf\overline{p}) = 0.49 & P(b\overline{f}p) = 0.045 & P(b\overline{f}\overline{p}) = 0.01 \\
P(\overline{b}fp) = 0.0 & P(\overline{b}f\overline{p}) = 0.2 & P(\overline{b}\overline{f}p) = 0.0 & P(\overline{b}\overline{f}\overline{p}) = 0.25
\end{array}
$$

it holds that $P \models^{pr} \mathcal{K}$ as e. g. $P(b) = P(bfp) + P(bf\overline{p}) + P(b\overline{f}p) + P(b\overline{f}\overline{p}) = 0.55$ and $P(bf) = P(bfp) + P(bf\overline{p}) = 0.495$ and therefore $P(f \mid b) = {}^{P(bf)}/{}_{P(b)} = 0.9$.

A probabilistic conditional $(\psi \mid \phi)[d]$ is *normal* if and only if there are $\omega, \omega' \in \Omega(\mathsf{At})$ with $\omega \models \psi\phi$ and $\omega' \models \overline{\psi}\phi$.[1] In other words, a probabilistic conditional c is normal if it is satisfiable but not tautological.

Example 2. The probabilistic conditionals $c_1 = (\top \mid a)[1]$ and $c_2 = (\overline{a} \mid a)[0.1]$ are *not* normal as c_1 is tautological (there is no $\omega \in \Omega(\mathsf{At})$ with $\omega \models \overline{\top}a$ as $\overline{\top}a \equiv \perp$) and c_2 is not satisfiable (there is no $\omega \in \Omega(\mathsf{At})$ with $\omega \models \overline{a}a$ as $\overline{a}a \equiv \perp$)

As a technical convenience, for the rest of this paper we consider only normal probabilistic conditionals, so let $\mathbb{K}$ be the set of all knowledge bases of $\mathcal{C}(\mathcal{L}(\mathsf{At}))$ that contain only normal probabilistic conditionals.

Proposition 1. *If $(\psi \mid \phi)[d]$ is normal then $(\psi \mid \phi)[x]$ is normal for every $x \in [0, 1]$.*

The proof of the above proposition is easy to see as the definition of normality does not depend on the probability of a conditional.

Knowledge bases $\mathcal{K}_1, \mathcal{K}_2$ are *extensionally equivalent*, denoted by $\mathcal{K}_1 \equiv^e \mathcal{K}_2$, if and only if $\mathsf{Mod}(\mathcal{K}_1) = \mathsf{Mod}(\mathcal{K}_2)$. Note that the notion of extensional equivalence does not distinguish between inconsistent knowledge bases, i. e. for inconsistent $\mathcal{K}_1$ and $\mathcal{K}_2$ it always holds that $\mathcal{K}_1 \equiv^e \mathcal{K}_2$. Consequently, we also consider another equivalence relation for knowledge bases. Knowledge bases $\mathcal{K}_1, \mathcal{K}_2$ are *semi-extensionally equivalent*, denoted by $\mathcal{K}_1 \equiv^s \mathcal{K}_2$, if and only if there is a bijection $\rho_{\mathcal{K}_1,\mathcal{K}_2} : \mathcal{K}_1 \to \mathcal{K}_2$ such that $c \equiv^e \rho_{\mathcal{K}_1,\mathcal{K}_2}(c)$ for every $c \in \mathcal{K}_1$. Note that $\mathcal{K}_1 \equiv^s \mathcal{K}_2$ implies $\mathcal{K}_1 \equiv^e \mathcal{K}_2$ but the other direction is not true in general.

Example 3. Consider the two knowledge bases $\mathcal{K}_1 = \langle (a)[0.7], (a)[0.4] \rangle$ and $\mathcal{K}_2 = \langle (b)[0.8], (b)[0.3] \rangle$. Both $\mathcal{K}_1$ and $\mathcal{K}_2$ are inconsistent and therefore $\mathcal{K}_1 \equiv^e \mathcal{K}_2$. But it holds that $\mathcal{K}_1 \not\equiv^s \mathcal{K}_2$ as both $(a)[0.7] \not\equiv^e (b)[0.8]$ and $(a)[0.7] \not\equiv^e (b)[0.3]$.

One way for reasoning with knowledge bases is by using model-based inductive reasoning techniques [6]. For example, reasoning based on the *principle of maximum entropy* selects among the models of a knowledge base $\mathcal{K}$ the one unique probability function with maximum entropy. Reasoning with this model satisfies several commonsense properties, see e. g. [6, 5]. However, a necessary requirement for the application of model-based inductive reasoning techniques is the existence of at least one model of a knowledge base. In order to reason with inconsistent knowledge bases the inconsistency has to be resolved first. In the following, we discuss the topic of *inconsistency measurement* for probabilistic conditional logic as inconsistency measures can support the knowledge engineer in the task of resolving inconsistency.

[1] I thank an anonymous reviewer for pointing this formalization out to me.

3 Principles for Inconsistency Measurement

An *inconsistency measure* $\mathcal{I}$ is a function that maps a (possibly inconsistent) knowledge base onto a positive real value, i.e. a function $\mathcal{I} : \mathbb{K} \to [0, \infty)$. The value $\mathcal{I}(\mathcal{K})$ for a knowledge base $\mathcal{K}$ is called the *inconsistency value* for $\mathcal{K}$ with respect to $\mathcal{I}$. Intuitively, we want $\mathcal{I}$ to be a function on knowledge bases that is monotonically increasing with the inconsistency in the knowledge base. If the knowledge base is consistent, $\mathcal{I}$ shall be minimal. In order to formalize this intuition we give a list of principles that should be satisfied by any reasonable inconsistency measure. For that we need some further notation.

Definition 3. *A set $\mathcal{M}$ is* minimal inconsistent *if $\mathcal{M}$ is inconsistent and every $\mathcal{M}' \subsetneq \mathcal{M}$ is consistent.*

Let $\mathsf{MI}(\mathcal{K})$ be the set of the minimal inconsistent subsets of $\mathcal{K}$.

Example 4. Consider the knowledge base $\mathcal{K} = \langle (a)[0.3], (b)[0.5], (a \wedge b)[0.7] \rangle$. Then the set of minimal inconsistent subsets of $\mathcal{K}$ is given via

$$\mathsf{MI}(\mathcal{K}) = \{ \ \{(a)[0.3], (a \wedge b)[0.7]\}, \ \{(b)[0.5], (a \wedge b)[0.7]\} \ \} \quad .$$

The notion of minimal inconsistent subsets captures those conditionals that are responsible for creating inconsistencies. Conditionals that do not take part in creating an inconsistency are *free*.

Definition 4. *A probabilistic conditional $c \in \mathcal{K}$ is* free *in $\mathcal{K}$ if and only if $c \notin \mathcal{M}$ for all $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$.*

For a conditional or a knowledge base C let $\mathsf{At}(C)$ denote the set of atoms appearing in C.

Definition 5. *A probabilistic conditional $c \in \mathcal{K}$ is* safe *in $\mathcal{K}$ if and only if $\mathsf{At}(c) \cap \mathsf{At}(\mathcal{K} \setminus c) = \emptyset$.*

Note that the notion of safeness is due to Hunter and Konieczny [4]. The notion of a free conditional is clearly more general than the notion of a safe conditional.

Proposition 2. *If c is safe in $\mathcal{K}$ then c is free in $\mathcal{K}$.*

The proof of Proposition 2 can be found in [11].

Definition 6. *Let $\mathcal{K} \in \mathbb{K}$ be a knowledge base with $\mathcal{K} = \langle c_1, \ldots, c_n \rangle$ and $c_i = (\psi_i \,|\, \phi_i)[d_i]$ for $i = 1, \ldots, n$. The function $\Lambda_\mathcal{K} : [0,1]^n \to \mathbb{K}$ with $\Lambda_\mathcal{K}(x_1, \ldots, x_n) = \langle (\psi_1 \,|\, \phi_1)[x_1], \ldots, (\psi_n \,|\, \phi_n)[x_n] \rangle$ is called the* characteristic function *of $\mathcal{K}$.*

Due to Proposition 1 the function $\Lambda_\mathcal{K}$ is well-defined. The above definition is also the justification for imposing an order on the probabilistic conditionals of a knowledge base.

Definition 7. *Let $\mathcal{I}$ be an inconsistency measure and let $\mathcal{K}$ be a knowledge base. The function $\theta_{\mathcal{I},\mathcal{K}} : [0,1]^{|\mathcal{K}|} \to [0,\infty)$ with $\theta_{\mathcal{I},\mathcal{K}} = \mathcal{I} \circ \Lambda_{\mathcal{K}}$ is called the* characteristic inconsistency function *of $\mathcal{I}$ and $\mathcal{K}$.*

Consider now the following properties from [10]. Let $\mathcal{K}, \mathcal{K}'$ be knowledge bases and c a probabilistic conditional.

Consistency. $\mathcal{K}$ is consistent if and only if $\mathcal{I}(\mathcal{K}) = 0$
Monotonicity. $\mathcal{I}(\mathcal{K}) \leq \mathcal{I}(\mathcal{K} \cup \{c\})$
Super-additivity. If $\mathcal{K} \cap \mathcal{K}' = \emptyset$ then $\mathcal{I}(\mathcal{K} \cup \mathcal{K}') \geq \mathcal{I}(\mathcal{K}) + \mathcal{I}(\mathcal{K}')$
Weak independence. If $c \in \mathcal{K}$ is safe in $\mathcal{K}$ then $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K} \setminus \{c\})$
Independence. If $c \in \mathcal{K}$ is free in $\mathcal{K}$ then $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K} \setminus \{c\})$
Penalty. If $c \in \mathcal{K}$ is not free in $\mathcal{K}$ then $\mathcal{I}(\mathcal{K}) > \mathcal{I}(\mathcal{K} \setminus \{c\})$
Continuity. $\theta_{\mathcal{I},\mathcal{K}}$ is continuous

The property *consistency* demands that $\mathcal{I}(\mathcal{K})$ is minimal for consistent $\mathcal{K}$. The properties *monotonicity* and *super-additivity* demand that $\mathcal{I}$ is non-decreasing under the addition of new information. The properties *weak independence* and *independence* say that the inconsistency value should stay the same when adding "harmless" information. The property *penalty* is the counterpart of *independence* and demands that adding inconsistent information increases the inconsistency value. The final property *continuity* describes our main demand for continuous inconsistency measurement, i. e., a "slight" change in the knowledge base should not result in a "vast" change of the inconsistency value.

We also consider the following novel properties. If f is a function $f : [0,1]^n \to [0,\infty)$ then $\nabla f : K \to \mathbb{R}^n$ with $\nabla f(x_1,\ldots,x_n) = (\partial f/\partial x_1,\ldots,\partial f/\partial x_n)$ is its gradient with partial derivatives $\partial f/\partial x_1,\ldots,\partial f/\partial x_n$. There, $K \subseteq [0,1]^n$ is the subset of the domain of f where f is differentiable with respect to all directions.

Irrelevance of syntax. If $\mathcal{K}_1 \equiv^s \mathcal{K}_2$ then $\mathcal{I}(\mathcal{K}_1) = \mathcal{I}(\mathcal{K}_2)$
MI-separability. If $\mathsf{MI}(\mathcal{K}_1 \cup \mathcal{K}_2) = \mathsf{MI}(\mathcal{K}_1) \cup \mathsf{MI}(\mathcal{K}_2)$ and $\mathsf{MI}(\mathcal{K}_1) \cap \mathsf{MI}(\mathcal{K}_2) = \emptyset$ then $\mathcal{I}(\mathcal{K}_1 \cup \mathcal{K}_2) = \mathcal{I}(\mathcal{K}_1) + \mathcal{I}(\mathcal{K}_2)$
Differentiability. $\theta_{\mathcal{I},\mathcal{K}}$ is differentiable in $(0,1)^{|\mathcal{K}|}$
Weak differentiability. $\theta_{\mathcal{I},\mathcal{K}}$ is differentiable almost everywhere in $(0,1)^{|\mathcal{K}|}$
Sub-linearity. $\mathsf{Im}\ \nabla\theta_{\mathcal{I},\mathcal{K}} \subseteq [-1,1]^{|\mathcal{K}|}$

We define the property *irrelevance of syntax* in terms of the equivalence relation $\equiv^s$ as all inconsistent knowledge bases are equivalent with respect to $\equiv^e$. For an inconsistency measure $\mathcal{I}$, imposing *irrelevance of syntax* to hold in terms of $\equiv^e$ would yield $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K}')$ for every two inconsistent knowledge bases $\mathcal{K}, \mathcal{K}'$. The property *MI-separability*—which has been adapted from [3]—states that determining the value of $\mathcal{I}(\mathcal{K}_1 \cup \mathcal{K}_2)$ can be split into determining the values of $\mathcal{I}(\mathcal{K}_1)$ and $\mathcal{I}(\mathcal{K}_2)$ if the minimal inconsistent subsets of $\mathcal{K}_1 \cup \mathcal{K}_2$ are partitioned by $\mathcal{K}_1$ and $\mathcal{K}_2$. The property *differentiability* strengthens the property *continuity* and expects $\mathcal{I}$ to behave even more smoothly. The property *weak differentiability* allows $\mathcal{I}$ to be non-differentiable on a null set. Finally, the property *sub-linearity* demands that the value $\mathcal{I}(\mathcal{K})$ changes at most linearly in the change of $\mathcal{K}$. This means, for example, that if one changes the probability of a conditional in $\mathcal{K}$ by some value α, then the difference between the corresponding values of $\mathcal{I}$ should not be more than α.

Some relationships between the above properties are as follows.

Proposition 3. *Let $\mathcal{I}$ be an inconsistency measure and let $\mathcal{K}, \mathcal{K}'$ be some knowledge bases.*

1. *If $\mathcal{I}$ satisfies* super-additivity *then $\mathcal{I}$ satisfies* monotonicity.
2. *If $\mathcal{I}$ satisfies* independence *then $\mathcal{I}$ satisfies* weak independence.
3. *If $\mathcal{I}$ satisfies* MI-separability *then $\mathcal{I}$ satisfies* independence.
4. *If $\mathcal{I}$ satisfies* differentiability *then $\mathcal{I}$ satisfies* continuity.
5. *If $\mathcal{I}$ satisfies* differentiability *then $\mathcal{I}$ satisfies* weak differentiability.
6. *$\mathcal{K} \subseteq \mathcal{K}'$ implies $\mathsf{MI}(\mathcal{K}) \subseteq \mathsf{MI}(\mathcal{K}')$.*
7. *If $\mathcal{I}$ satisfies* independence *then $\mathsf{MI}(\mathcal{K}) = \mathsf{MI}(\mathcal{K}')$ implies $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K}')$.*
8. *If $\mathcal{I}$ satisfies* independence *and* penalty *then $\mathsf{MI}(\mathcal{K}) \subsetneq \mathsf{MI}(\mathcal{K}')$ implies $\mathcal{I}(\mathcal{K}) < \mathcal{I}(\mathcal{K}')$.*

The proofs of 1.)-3.) and 6.)-8.) can be found in [11]. The proofs of 4.) and 5.) are obvious.

Previous research on inconsistency measurement focuses on inconsistency measurement on propositional logic, see e. g. [3]. Adopting those measures for probabilistic conditional logic is straightforward [11]. For example, consider the following definition.

Definition 8. *The function $\mathcal{I}^{\#} : \mathbb{K} \to [0, \infty)$ defined via $\mathcal{I}^{\#}(\mathcal{K}) = |\mathsf{MI}(\mathcal{K})|$ is called the* MI cardinality measure.

The MI cardinality measure determines the inconsistency value of a knowledge base $\mathcal{K}$ as the number of minimal inconsistent subsets of $\mathcal{K}$.

Example 5. We continue Ex. 4. There it holds that $\mathcal{I}^{\#}(\mathcal{K}) = 2$.

Although $\mathcal{I}^{\#}$ is a rather simple inconsistency measure it already complies with many principles.

Proposition 4. *The function $\mathcal{I}^{\#}$ satisfies* consistency, monotonicity, super-additivity, weak independence, independence, MI-separability, *and* penalty.

The proof of Proposition 4 can be found in [11]. However, as the following example shows, the MI cardinality measure—and other inconsistency measures that were developed for propositional logic—does not satisfy *continuity* which is a major drawback for the probabilistic setting.

Example 6. Consider the knowledge base $\mathcal{K} = \langle (b \,|\, a)[1], (a)[1], (b)[0] \rangle$ which models strongly inconsistent information. Clearly, it holds that $\mathcal{I}^{\#}(\mathcal{K}) = 1$. Consider now the two modifications $\mathcal{K}', \mathcal{K}''$ of $\mathcal{K}$ given via

$$\mathcal{K}' = \langle (b \,|\, a)[0.6], (a)[0.6], (b)[0.3599] \rangle$$
$$\mathcal{K}'' = \langle (b \,|\, a)[0.6], (a)[0.6], (b)[0.36] \rangle \quad .$$

It is also clear that $\mathcal{I}^{\#}(\mathcal{K}') = 1$ and $\mathcal{I}^{\#}(\mathcal{K}'') = 0$. By comparing $\mathcal{K}'$ and $\mathcal{K}''$ one can discover only a minor difference of the modeled knowledge. From a practical point of view, whether b has probability 0.3599 or 0.36 may not matter for the intended application. Still, a knowledge engineer may not grasp the harmlessness of the inconsistency in $\mathcal{K}'$ as $\mathcal{K}'$ and $\mathcal{K}$ have the same inconsistency value.

In the following, we discuss inconsistency measures that are more apt for the probabilistic setting.

4 Measuring Inconsistency by Distance Minimization

As can be seen in Ex. 6 the probabilities of conditionals play a crucial role in creating inconsistencies. In order to respect this role we propose a family of inconsistency measures that is based on the distance to consistency. Afterwards we propose a compound measure that uses this measure and behaves well with the desired properties.

Before defining the measure we need some further notation. Knowledge bases $\mathcal{K}_1, \mathcal{K}_2$ are *qualitatively equivalent*, denoted by $\mathcal{K}_1 \cong^q \mathcal{K}_2$, if and only if there is a bijection $\sigma_{\mathcal{K}_1,\mathcal{K}_2} : \mathcal{K}_1 \to \mathcal{K}_2$ such that $\mathsf{body}(c) \equiv \mathsf{body}(\sigma(c))$ and $\mathsf{head}(c) \wedge \mathsf{body}(c) \equiv \mathsf{head}(\sigma(c)) \wedge \mathsf{body}(\sigma(c))$ for every $c \in \mathcal{K}_1$. Note that the function $\sigma_{\mathcal{K}_1,\mathcal{K}_2}$ might not be uniquely determined.

Example 7. Consider the knowledge bases $\mathcal{K}_1 = \langle (a)[0.2], (a)[0.8] \rangle$ and $\mathcal{K}_2 = \langle (a)[0.3], (a)[0.9] \rangle$. It holds that $\mathcal{K}_1 \cong^q \mathcal{K}_2$ but there are two bijections $\sigma^1_{\mathcal{K}_1,\mathcal{K}_2}$ and $\sigma^2_{\mathcal{K}_1,\mathcal{K}_2}$ given via

$$\sigma^1_{\mathcal{K}_1,\mathcal{K}_2}((a)[0.2]) = (a)[0.3] \qquad \sigma^1_{\mathcal{K}_1,\mathcal{K}_2}((a)[0.8]) = (a)[0.9]$$
$$\sigma^2_{\mathcal{K}_1,\mathcal{K}_2}((a)[0.2]) = (a)[0.9] \qquad \sigma^1_{\mathcal{K}_1,\mathcal{K}_2}((a)[0.8]) = (a)[0.3]$$

that establish the qualitative equivalence of $\mathcal{K}_1$ and $\mathcal{K}_2$.

If $\mathcal{K}_1 \cong^q \mathcal{K}_2$ let $\mathcal{S}_{\mathcal{K}_1,\mathcal{K}_2}$ be the set of bijections between $\mathcal{K}_1$ and $\mathcal{K}_2$ with the above property. Note that $\mathcal{S}_{\mathcal{K}_1,\mathcal{K}_2}$ is finite as both $\mathcal{K}_1$ and $\mathcal{K}_2$ are finite. Let $\mathbb{N}^+$ denote the set of positive integers.

Definition 9. *Let $\mathcal{K}_1, \mathcal{K}_2$ be some knowledge bases and let $p \in \mathbb{N}^+$. Then the* p-norm distance $d^p(\mathcal{K}_1, \mathcal{K}_2)$ *of $\mathcal{K}_1$ to $\mathcal{K}_2$ is defined via*

$$d^p(\mathcal{K}_1, \mathcal{K}_2) = \begin{cases} \min\limits_{\sigma \in \mathcal{S}_{\mathcal{K}_1,\mathcal{K}_2}} \left\{ \sqrt[p]{\sum_{c \in \mathcal{K}_1} |\mathsf{prob}(c) - \mathsf{prob}(\sigma(c))|^p} \right\} & \textit{if } \mathcal{K}_1 \cong^q \mathcal{K}_2 \\ \infty & \textit{otherwise} \end{cases} .$$

Note that d^p is indeed a distance measure, i. e., it is positive definite, symmetric, and satisfies the triangle inequality. This measure assigns an infinite distance to two knowledge bases $\mathcal{K}_1, \mathcal{K}_2$ iff $\mathcal{K}_1, \mathcal{K}_2$ are not qualitatively equivalent. Otherwise it is equivalent to the standard p-norm distance by interpreting probabilities of conditionals as coordinates and selecting a bijection $\sigma \in \mathcal{S}_{\mathcal{K}_1,\mathcal{K}_2}$ that minimizes this distance.

Example 8. For $\mathcal{K}_1$ and $\mathcal{K}_2$ as given in Ex. 7 it holds that $d^1(\mathcal{K}_1, \mathcal{K}_2) = 0.2$ and $d^2(\mathcal{K}_1, \mathcal{K}_2) \approx 0.1414$. Note that $\sigma^1_{\mathcal{K}_1,\mathcal{K}_2}$ is used for determining $d^p(\mathcal{K}_1, \mathcal{K}_2)$ as $\sigma^2_{\mathcal{K}_1,\mathcal{K}_2}$ yields values 1.2 and ≈ 0.8602, respectively.

The following definition has been rephrased from [10, 7].

Definition 10. *Let $\mathcal{K}$ be a knowledge base and let $p \in \mathbb{N}^+$. Then define the* d^p-measure $\mathcal{I}^p$ *via*

$$\mathcal{I}^p(\mathcal{K}) = \min\{d^p(\mathcal{K}, \mathcal{K}') \mid \mathcal{K}' \textit{ consistent}\} \tag{1}$$

for a knowledge base $\mathcal{K}$.

The value $\mathcal{I}^p(\mathcal{K})$ is the minimal distance to a knowledge base $\mathcal{K}'$ that is both qualitatively equivalent to $\mathcal{K}$ and consistent. Now we can also justify representing knowledge bases as multi-sets. Considering the knowledge base $\mathcal{K} = \langle (a)[0.2], (a)[0.6] \rangle$, it holds that $\mathcal{K}' = \langle (a)[0.4], (a)[0.4] \rangle$ minimizes the p-norm distance to $\mathcal{K}$.

The above definition presupposes that the minimum in Equation (1) exists. The following proposition shows that this is indeed the case.

Proposition 5. *The function $\mathcal{I}^p$ is well-defined.*

Proof. *Let $\mathcal{K} = \langle (\psi_1 \mid \phi_1)[d_1], \ldots, (\psi_n \mid \phi_n)[d_n] \rangle$ be a knowledge base and let P_0 be the uniform probability function on $\Omega(\mathsf{At})$, i. e, it holds that $P_0(\omega) = 1/|\Omega(\mathsf{At})|$ for every $\omega \in \Omega(\mathsf{At})$ (note that $\Omega(\mathsf{At})$ is finite as At is finite). Let $\mathcal{K}'$ be the knowledge base defined via*

$$\mathcal{K}' = \langle (\psi_1 \mid \phi_1)[P_0(\psi_1 \mid \phi_1)], \ldots, (\psi_n \mid \phi_n)[P_0(\psi_n \mid \phi_n)] \rangle$$

As P_0 is a positive probability function and every $c \in \mathcal{K}$ is normal it follows that $\mathcal{K}'$ is well-defined and $P_0 \models^{pr} \mathcal{K}'$. As $\mathcal{K} \cong^q \mathcal{K}'$ it follows that $\mathcal{I}^p(\mathcal{K})$ is finite. Furthermore, observe that the set

$$D_{\mathcal{K}} = \{\langle x_1, \ldots, x_n \rangle \mid \langle (\psi_1 \mid \phi_1)[x_1], \ldots, (\psi_n \mid \phi_n)[x_n] \rangle \textit{ is consistent }\}$$

is compact (bounded and closed) as probabilistic satisfaction is defined via the equation $P(\psi\phi) = dP(\phi)$ (for a probabilistic conditional $(\psi \mid \phi)[d]$). As the functional mapping

$$\langle x_1, \ldots, x_n \rangle \mapsto \sqrt[p]{|d_1 - x_1|^p + \ldots + |d_n - x_n|^p}$$

is continuous it follows that the set $\{d^p(\mathcal{K}, \mathcal{K}') \mid \mathcal{K}' \textit{ consistent}\}$ is closed. Therefore, the minimum of this set and the value of $\mathcal{I}^p$ is well-defined. $\square$

In [7] it has been shown that for every $p, p' \in \mathbb{N}^+$ with $p \neq p'$ the two measures $\mathcal{I}^p$ and $\mathcal{I}^{p'}$ are not equivalent, i. e., there are knowledge bases $\mathcal{K}_1$ and $\mathcal{K}_2$ such that $\mathcal{I}^p(\mathcal{K}_1) > \mathcal{I}^p(\mathcal{K}_2)$ but $\mathcal{I}^{p'}(\mathcal{K}_1) < \mathcal{I}^{p'}(\mathcal{K}_2)$.

Example 9. We continue Ex. 6. There, the knowledge base $\mathcal{K}^* = \langle (b \,|\, a)[1], (a)[0.5], (b)[0.5] \rangle$ satisfies $\mathcal{I}^p(\mathcal{K}) = d^p(\mathcal{K}, \mathcal{K}^*)$ for every p. In particular, it holds that $\mathcal{I}^p(\mathcal{K}) = \sqrt[p]{2 \cdot 0.5^p}$. For example, it holds that $\mathcal{I}^1(\mathcal{K}) = 1$ and $\mathcal{I}^2(\mathcal{K}) \approx 0.707$. Furthermore, it holds that $\mathcal{I}^1(\mathcal{K}') = 0.0001$ and $\mathcal{I}^2(\mathcal{K}') \approx 0.00006$, and clearly $\mathcal{I}^1(\mathcal{K}'') = \mathcal{I}^2(\mathcal{K}'') = 0$.

We also propose the following compound measure that explicitly considers the crucial role of minimal inconsistent subsets.

Definition 11. *Let $\mathcal{K}$ be a knowledge base and let $\mathcal{I}$ be an inconsistency measure. Then define the* MI-measure $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K})$ *of $\mathcal{K}$ and $\mathcal{I}$ via*

$$\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K}) = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \mathcal{I}(\mathcal{M}) \quad .$$

The MI-measure is defined as the sum of the inconsistency values of all minimal inconsistent subsets of the knowledge base under consideration. In the next section, we investigate the properties of the measures proposed above.

5 Analysis and Comparison

We first investigate the properties of the d^p-measure. We can extend a result from [10] as follows.

Theorem 1. *If $p \in \mathbb{N}^+$ then $\mathcal{I}^p$ satisfies* consistency*,* monotonicity*,* weak independence*,* independence*,* irrelevance of syntax*,* continuity*,* weak differentiability*, and* sub-linearity*.*

Proof.

Consistency*. As $\mathcal{K}$ is consistent and $d^p(\mathcal{K}, \mathcal{K}) = 0$ it follows directly $\mathcal{I}^p(\mathcal{K}) = 0$.*

Monotonicity*. Let $\mathcal{K} = \langle c_1, \ldots, c_n \rangle$ and let $\mathcal{K}'$ be consistent and $d^p(\mathcal{K}, \mathcal{K}') = \mathcal{I}^p(\mathcal{K})$. Let furthermore $\sigma_{\mathcal{K},\mathcal{K}'} \in \mathcal{S}_{\mathcal{K}_1,\mathcal{K}_2}$ be the bijection used to determine $d^p(\mathcal{K}, \mathcal{K}')$. It follows that $\mathcal{K}'' = \mathcal{K}' \setminus \{\sigma_{\mathcal{K},\mathcal{K}'}(c_n)\}$ is consistent as well and $\mathcal{K} \setminus \{c_n\} \cong^q \mathcal{K}''$. It follows that $\mathcal{I}^p(\mathcal{K} \setminus \{c_n\}) \leq d^p(\mathcal{K} \setminus \{c_n\}, \mathcal{K}'')$. Setting $a_i = |\mathsf{prob}(c_i) - \mathsf{prob}(\sigma_{\mathcal{K},\mathcal{K}'}(c_i))|$ for $i = 1, \ldots, n$ we get*

$$\begin{aligned} \mathcal{I}^p(\mathcal{K}) = d^p(\mathcal{K}, \mathcal{K}') &= \sqrt[p]{a_1^p + \ldots + a_n^p} \\ &\geq \sqrt[p]{a_1^p + \ldots + a_{n-1}^p} = d^p(\mathcal{K} \setminus \{c\}, \mathcal{K}'') \geq \mathcal{I}^p(\mathcal{K} \setminus \{c_n\}) \end{aligned}$$

Independence*. In [11] it has been shown that $\mathcal{I}^p$ for $p = 1$ satisfies* independence*. This result can be extended to arbitrary p in a straightforward fashion.*

Irrelevance of syntax*. Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be knowledge bases with $\mathcal{K}_1 \equiv^s \mathcal{K}_2$. Let $\mathcal{K}_1'$ be consistent such that $\mathcal{I}^p(\mathcal{K}_1) = d^p(\mathcal{K}_1, \mathcal{K}_1')$. It follows that $\mathcal{K}_1 \cong^q \mathcal{K}_1'$. As $\mathcal{K}_1 \equiv^s \mathcal{K}_2$ there is a consistent $\mathcal{K}_2'$ such that $\mathcal{K}_1' \equiv^s \mathcal{K}_2'$ and $\mathcal{K}_2 \cong^q \mathcal{K}_2'$. It follows that $\mathcal{I}^p(\mathcal{K}_2) \leq d^p(\mathcal{K}_2, \mathcal{K}_2') = d^p(\mathcal{K}_1, \mathcal{K}_1') = \mathcal{I}^p(\mathcal{K}_1)$. Similarly we obtain $\mathcal{I}^p(\mathcal{K}_1) \leq \mathcal{I}^p(\mathcal{K}_2)$ and therefore the claim.*

Continuity. *In [11] it has been shown that* $\mathcal{I}^p$ *for* $p = 1$ *satisfies* continuity. *This result can be extended to arbitrary* p *in a straightforward fashion.*

Weak differentiability. *We only give a proof sketch for* weak differentiability. *Let* $\vec{x} \in (0,1)^{|\mathcal{K}|}$ *such that there is an open* ϵ*-ball* B_ϵ *with* $\vec{x} \in B_\epsilon$ *and* $\theta_{\mathcal{I}^p,\mathcal{K}}(\vec{y}) > 0$ *for every* $\vec{y} \in B_\epsilon$. *Then* $\theta_{\mathcal{I}^p,\mathcal{K}}$ *is differentiable on* B_ϵ *as the p-norm distance is a differentiable function if the distance does not equal zero. Furthermore, let now* $\vec{x} \in (0,1)^{|\mathcal{K}|}$ *be such that there is an open* ϵ*-ball* B_ϵ *with* $\vec{x} \in B_\epsilon$ *and* $\theta_{\mathcal{I}^p,\mathcal{K}}(\vec{y}) = 0$ *for every* $\vec{y} \in B_\epsilon$. *Then* $\theta_{\mathcal{I}^p,\mathcal{K}}$ *is differentiable on* B_ϵ *as it is a constant function. Note furthermore that the set* $C \subseteq (0,1)^{|\mathcal{K}|}$ *such that for every* $\vec{y} \in C$ *it holds that* $\theta_{\mathcal{I}^p,\mathcal{K}}(\vec{y}) = 0$ *is the finite union of pair-wise disjoint closed convex sets* $F_1, \ldots, F_m$*, cf. [11]. Without loss of generality, let* $F_1, \ldots, F_k$ *with* $k \leq m$ *be the sets with dimension* $|\mathcal{K}|$ *and* $F_{k+1}, \ldots, F_m$ *be the sets with a dimension less than* $|\mathcal{K}|$. *Let* $\mathsf{bd}\ S$ *denote the boundary of a set* S. *Note that* $\mathsf{bd}\ F_i$ *has dimension* $|\mathcal{K}| - 1$ *for* $i = 1, \ldots, k$. *Then the set* $F = \mathsf{bd}\ F_1 \cup \ldots \mathsf{bd}\ F_k \cup F_{k+1} \cup \ldots F_m$ *is a null set in* $(0,1)^{|\mathcal{K}|}$ *and* $\theta_{\mathcal{I}^p,\mathcal{K}}$ *is differentiable on* $(0,1)^{|\mathcal{K}|} \setminus F$.

Sub-linearity. *We only give a proof sketch for* sub-linearity. *Let* $\mathcal{K}$ *be the knowledge base* $\mathcal{K} = \langle (\psi_1 \,|\, \phi_1)[d_1], \ldots, (\psi_n \,|\, \phi_n)[d_n] \rangle$ *and let* $\vec{x} \in [0,1]^{|\mathcal{K}|}$ *such that* $\theta_{\mathcal{I},\mathcal{K}}$ *is differentiable in* $\vec{x} = (x_1, \ldots, x_n)$. *Note that*

$$\left(\sqrt[p]{g(x)}\right)' = \frac{1}{p}\frac{1}{g(x)^{p-1}}g'(x)$$

for differentiable g *and that* $|(|x|)'| = 1$ *for* $x \neq 0$. *Then consider the function*

$$f(\vec{x}) = \sqrt[p]{|d_1 - x_1|^p + \ldots + |d_n - x_n|^p} \tag{2}$$

and the following bound on the absolute value of its partial derivatives ($i = 1, \ldots, n$*)*

$$\begin{aligned}
\left|\frac{\partial f}{\partial x_i}\right| &= \left|\frac{1}{p}\frac{1}{\left(\sqrt[p]{|d_1 - x_1|^p + \ldots + |d_n - x_n|^p}\right)^{p-1}} \cdot p \cdot |d_i - x_i|^{p-1}\right| \\
&= \left|\left(\frac{|d_i - x_i|}{\sqrt[p]{|d_1 - x_1|^p + \ldots + |d_n - x_n|^p}}\right)^{p-1}\right| \\
&= \left|\left(\sqrt[p]{\frac{|d_i - x_i|^p}{|d_1 - x_1|^p + \ldots + |d_n - x_n|^p}}\right)^{p-1}\right| \\
&\leq 1
\end{aligned}$$

The above means that $d^p(\mathcal{K}, \mathcal{K}')$ *is sub-linear in* $\mathcal{K}'$ *for fixed* $\mathcal{K}$. *Assume now that there is an open* ϵ*-ball* B_ϵ *with* $\vec{x} \in B_\epsilon$ *and* $\theta_{\mathcal{I}^p,\mathcal{K}}(\vec{y}) > 0$ *for every* $\vec{y} \in B_\epsilon$. *Then* $|\partial\theta_{\mathcal{I}^p,\mathcal{K}}/\partial x_i| \leq 1$ *directly from above (as* $\theta_{\mathcal{I}^p,\mathcal{K}}$ *behaves like* f *in the worst case). Furthermore, let now* $\vec{x} \in (0,1)^{|\mathcal{K}|}$ *be such that there is an open* ϵ*-ball* B_ϵ *with* $\vec{x} \in B_\epsilon$ *and* $\theta_{\mathcal{I}^p,\mathcal{K}}(\vec{y}) = 0$ *for every* $\vec{y} \in B_\epsilon$. *Then clearly* $|\partial\theta_{\mathcal{I}^p,\mathcal{K}}/\partial x_i| = 0 \leq 1$.

Due to Proposition 3 it also follows that $\mathcal{I}^p$ satisfies weak independence. $\square$

Note that $\mathcal{I}^p$ does not satisfy *differentiability* in general as the following example shows.

Example 10. Consider the knowledge base $\mathcal{K} = \langle (a)[0.7], (a)[0.3] \rangle$. It is easy to see that e. g. $\theta_{\mathcal{I}^1,\mathcal{K}}(x,y) = |x-y|$. In particular, it holds that $\theta_{\mathcal{I}^1,\mathcal{K}}(x,y) = 0$ if and only if $x = y$. It also also quite clear that the absolute value $|x|$ is continuous for all x but only differentiable for $x \neq 0$.

However, for $p > 1$ we can strengthen Theorem 1 as follows.

Theorem 2. *If $p \in \mathbb{N}^+$ and $p > 1$ then $\mathcal{I}^p$ satisfies* differentiability.

We omit the proof of the above theorem due to space restrictions but note that this follows from the differentiability of the p-norm distance for $p > 1$. Observe that $\mathcal{I}^p$ does not satisfy *penalty* which has been mistakenly claimed in [10]. Consider the following counterexample.

Example 11. Consider the knowledge base $\mathcal{K} = \langle (a)[0.7], (a)[0.3] \rangle$ and the probabilistic conditional $(a)[0.5]$. Then $(a)[0.5]$ is not free in $\mathcal{K}' = \mathcal{K} \cup \{(a)[0.5]\}$ as $\{(a)[0.3], (a)[0.5]\} \in \mathsf{MI}(\mathcal{K}')$. However, it holds that $\mathcal{I}^1(\mathcal{K}) = \mathcal{I}^1(\mathcal{K}') = 0.4$—as $\langle (a)[0.5], (a)[0.5] \rangle$ has minimal distance to $\mathcal{K}$ and $\langle (a)[0.5], (a)[0.5], (a)[0.5] \rangle$ has minimal distance to $\mathcal{K}'$—which violates *penalty*.

In [11] it has been show that $\mathcal{I}^p$ for $p = 1$ also satisfies *MI-separability* and *super-additivity*. This is not true for arbitrary values of p as the following example shows.

Example 12. Let $\mathcal{K} = \langle (a)[0.3], (a)[0.7], (b)[0.3], (b)[0.7] \rangle$. It is easy to see that $\mathcal{I}^2(\mathcal{K}) = \sqrt{0.2^2 + 0.2^2 + 0.2^2 + 0.2^2} = 0.4$. It also holds that

$$\mathcal{I}^2(\langle (a)[0.3], (a)[0.7] \rangle) = \mathcal{I}^2(\langle (b)[0.3], (b)[0.7] \rangle) = \sqrt{0.2^2 + 0.2^2} \approx 0.283 \quad .$$

It follows that

$$\mathcal{I}^2(\mathcal{K}) < \mathcal{I}^2(\langle (a)[0.3], (a)[0.7] \rangle) + \mathcal{I}^2(\langle (b)[0.3], (b)[0.7] \rangle)$$

violating *super-additivity* and *MI-separability* as $\langle (a)[0.3], (a)[0.7] \rangle$ and $\langle (b)[0.3], (b)[0.7] \rangle$ partition the set of minimal inconsistent subsets of $\mathcal{K}$.

We now have a look at the properties of the MI-measure.

Theorem 3. *Let $\mathcal{I}$ be an inconsistency measure.*

1. *$\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* monotonicity, super-additivity, weak independence, independence, *and* MI-separability.
2. *If $\mathcal{I}$ satisfies* consistency *then $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* consistency *and* penalty.
3. *If $\mathcal{I}$ satisfies* irrelevance of syntax *then $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* irrelevance of syntax.
4. *If $\mathcal{I}$ satisfies* continuity *then $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* continuity.

5. *If $\mathcal{I}$ satisfies* differentiability *then $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* differentiability.
6. *If $\mathcal{I}$ satisfies* weak differentiability *then $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* weak differentiability.

Proof.

1. *We first show that $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* super-additivity*. If $\mathcal{K} \cap \mathcal{K}' = \emptyset$ then it holds that $\mathsf{MI}(\mathcal{K}) \cap \mathsf{MI}(\mathcal{K}') = \emptyset$ as well. Due to 6.) in Proposition 3 it follows that $\mathsf{MI}(\mathcal{K}) \cup \mathsf{MI}(\mathcal{K}') \subseteq \mathsf{MI}(\mathcal{K} \cup \mathcal{K}')$. It follows*

$$\begin{aligned}\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K} \cup \mathcal{K}') &= \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K} \cup \mathcal{K}')} \mathcal{I}(\mathcal{M}) \geq \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \mathcal{I}(\mathcal{M}) + \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K}')} \mathcal{I}(\mathcal{M}) \\ &= \mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K}) + \mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K}') \quad .\end{aligned}$$

   Due to 1.) in Proposition 3 it also follows that $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies monotonicity*. We now show that $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* MI-separability*. Let $\mathsf{MI}(\mathcal{K} \cup \mathcal{K}') = \mathsf{MI}(\mathcal{K}) \cup \mathsf{MI}(\mathcal{K}')$ and $\mathsf{MI}(\mathcal{K}) \cap \mathsf{MI}(\mathcal{K}') = \emptyset$. Then clearly*

$$\begin{aligned}\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K} \cup \mathcal{K}') &= \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K} \cup \mathcal{K}')} \mathcal{I}(\mathcal{M}) = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \mathcal{I}(\mathcal{M}) + \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K}')} \mathcal{I}(\mathcal{M}) \\ &= \mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K}) + \mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K}') \quad .\end{aligned}$$

   Due to 2.) and 3.) in Proposition 3 it also follows that $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies independence *and* weak independence.
2. *We first show that $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* consistency*. If $\mathcal{K}$ is consistent then $\mathsf{MI}(\mathcal{K}) = \emptyset$ and $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K}) = 0$. If $\mathcal{K}$ is inconsistent then there is a $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$ and as $\mathcal{I}$ satisfies* consistency *it follows that $\mathcal{I}(\mathcal{M}) > 0$. Hence, $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K}) > 0$ as well. We now show that $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}$ satisfies* penalty*. Let $c \in \mathcal{K}$ be a probabilistic conditional that is not free in $\mathcal{K}$. Due to 6.) in Proposition 3 it follows that $\mathsf{MI}(\mathcal{K} \setminus \{c\}) \subseteq \mathsf{MI}(\mathcal{K})$. As $c \notin \mathcal{K} \setminus \{c\}$ and there is at least one $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$ with $c \in \mathcal{M}$ it follows that $\mathsf{MI}(\mathcal{K} \setminus \{c\}) \subsetneq \mathsf{MI}(\mathcal{K})$. As $\mathcal{I}$ satisfies* consistency *it follows that $\mathcal{I}(\mathcal{M}) > 0$ and therefore $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K} \setminus \{c\}) < \mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K})$.*
3. *Let it hold that $\mathcal{K}_1 \equiv^s \mathcal{K}_2$. It follows that for every $\mathcal{M} \in \mathsf{MI}(\mathcal{K}_1)$ there is $\mathcal{M}' \in \mathsf{MI}(\mathcal{K}_2)$ with $\mathcal{M} \equiv^s \mathcal{M}'$, and vice versa. As $\mathcal{I}$ satisfies* irrelevance of syntax *it follows that $\mathcal{I}(\mathcal{M}) = \mathcal{I}(\mathcal{M}')$ for every $\mathcal{M} \in \mathsf{MI}(\mathcal{K}_1)$. Hence, it holds that $\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K}_1) = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K}_1)} \mathcal{I}(\mathcal{M}') = \sum_{\mathcal{M}' \in \mathsf{MI}(\mathcal{K}_2)} \mathcal{I}(\mathcal{M}') = \mathcal{I}^{\mathcal{I}}_{\mathsf{MI}}(\mathcal{K}_2)$.*
4. *It is easy to see that $\theta_{\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}},\mathcal{K}}$ is given via $\theta_{\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}},\mathcal{K}} = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \theta_{\mathcal{I},\mathcal{M}}$ (given an adequate ordering of the conditionals in $\mathcal{K}$). It follows directly, that $\theta_{\mathcal{I}^{\mathcal{I}}_{\mathsf{MI}},\mathcal{K}}$ is continuous if $\theta_{\mathcal{I},\mathcal{M}}$ is continuous for every $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$, i. e., if $\mathcal{I}$ satisfies* continuity.
5. *This holds due to the same argument used in 4.).*
6. *This holds due to the same argument used in 4.).* □

As one can see the MI-measure behaves very well with respect to our rationality postulates and even satisfies *penalty*, provided that the inner measure satisfies *consistency*.

Example 13. We continue Ex. 11. There it is $\mathcal{I}^{\mathcal{I}^1}_{\mathsf{MI}}(\mathcal{K}) = \mathcal{I}^p(\mathcal{K}) = 0.4$ but

$$\begin{aligned}\mathcal{I}^{\mathcal{I}^1}_{\mathsf{MI}}(\mathcal{K}') = \mathcal{I}^1(\langle (a)[0.7], (a)[0.3]\rangle) + \mathcal{I}^1(\langle (a)[0.7], (a)[0.5]\rangle)\\ + \mathcal{I}^1(\langle (a)[0.3], (a)[0.5]\rangle) = 0.4 + 0.2 + 0.2 = 0.8 \quad .\end{aligned}$$

Therefore, the addition of the conditional $(a)[0.5]$ is penalized by $\mathcal{I}^{\mathcal{I}^1}_{\mathsf{MI}}$.

The following corollary is a direct application of Theorems 1 and 3.

Corollary 1. *If $p \in \mathbb{N}^+$ then $\mathcal{I}^{\mathcal{I}^p}_{\mathsf{MI}}$ satisfies* consistency*,* monotonicity*,* super-additivity*,* weak independence*,* independence*,* MI-separability*,* penalty*,* irrelevance of syntax*,* continuity*, and* weak differentiability*. If $p > 1$ then $\mathcal{I}^{\mathcal{I}^p}_{\mathsf{MI}}$ also satisfies* differentiability*.*

Note that $\mathcal{I}^{\mathcal{I}^p}_{\mathsf{MI}}$ does not satisfy *sub-linearity* in general. Consider the following counterexample.

Example 14. Consider the knowledge base $\mathcal{K} = \langle (a)[0.7], (a)[0.3], (\neg a)[0.7]\rangle$. Note that for $x, y, z \in [0, 1]$ there are three (potential) minimal inconsistent subsets of $\Lambda_{\mathcal{K}}(x, y, z)$: $\{(a)[x], (a)[y]\}$, $\{(a)[x], (\neg a)[z]\}$, $\{(a)[y], (\neg a)[z]\}$. Then $\theta_{\mathcal{I}^1,\mathcal{K}}$ amounts to $\theta_{\mathcal{I}^1,\mathcal{K}}(x, y, z) = |x - y| + |1 - x - z| + |1 - y - z|$. For $x = y = 0$ and $z = 1$ we get $\theta_{\mathcal{I}^1,\mathcal{K}}(x, y, z) = 0$ and for $x = y = z = 0$ we get $\theta_{\mathcal{I}^1,\mathcal{K}}(x, y, z) = 2$. It follows that the absolute value of the partial derivation of $\theta_{\mathcal{I}^1,\mathcal{K}}$ with respect to the third coordinate has to be larger than 1 for at least one point.

6 Related Work

The work reported in this paper is based on results from [10, 7]. We extended the investigation of measuring inconsistency from [10] by introducing several novel rationality postulates, the MI-measure, and the resulting technical discussion. The d^p-measure has been proposed initially in [10] for $p = 1$ and extended to arbitrary values for p in [7]. The work [7] also contains an in-depth discussion of the d^p measure in terms of (among others) applicability and computability. The work [7] also defines probabilistic satisfaction via $P(\psi \mid \phi) = d$ which requires a more careful treatment of the case $P(\phi) = 0$ and the necessity of introducing infinitesimal inconsistency values. However, in [7] no evaluation of the d^p-measure in terms of rationality postulates is given.

The work [1] also investigates the problem of reasoning in inconsistent probabilistic knowledge bases. There, reasoning based on the principle of maximum entropy is extended to be applicable on inconsistent knowledge bases. By doing so one eliminates the need for restoring consistency. Furthermore, [1] also proposes a continuous inconsistency measure which rests on the notion of *candidacy functions*, a "fuzzy" extension of probability functions. A thorough comparison of the measure of [1] with our approach is outside the scope of this paper but we refer to [11] for a comparison with the d^1-measure. However, note that the measure of [1] does not satisfy *super-additivity*.

In [9] another continuous inconsistency measure for probabilistic conditional logic is proposed that is not based on the p-norm distance but on *generalized divergence* which is a specific distance for probability functions. However, no technical results and no evaluation is given in [9].

7 Summary

In this paper we investigated continuous inconsistency measures for probabilistic conditional logic. We built on previous work and introduced several novel rationality postulates for inconsistency measurement that addressed the behavior of inconsistency measures with respect to continuity. It turned out that our measures satisfy most of the desired properties and, in particular, the compound measure also satisfies *penalty*.

The d^1-measure has already been implemented within the Tweety library for artificial intelligence[2] and future work includes implementation of the other measures. This will enable us to evaluate the behavior of the measures in more depth.

References

1. Daniel, L.: Paraconsistent Probabilistic Reasoning. Ph.D. thesis, L'École Nationale Supérieure des Mines de Paris (2009)
2. Hansson, S.O.: A Textbook of Belief Dynamics: Theory Change and Database Updating. Springer-Verlag (1999)
3. Hunter, A., Konieczny, S.: Measuring Inconsistency through Minimal Inconsistent Sets. In: Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning. pp. 358–366. AAAI Press (2008)
4. Hunter, A., Konieczny, S.: On the Measure of Conflicts: Shapley Inconsistency Values. Artificial Intelligence 174(14), 1007–1026 (2010)
5. Kern-Isberner, G.: Conditionals in Nonmonotonic Reasoning and Belief Revision. No. 2087 in Lecture Notes in Computer Science, Springer-Verlag (2001)
6. Paris, J.B.: The Uncertain Reasoner's Companion – A Mathematical Perspective. Cambridge University Press (1994)
7. Picado-Muiño, D.: Measuring and Repairing Inconsistency in Probabilistic Knowledge Bases. International Journal of Approximate Reasoning (2011), to appear
8. Reiter, R.: A Logic for Default Reasoning. Artificial Intelligence 13(1–2), 81–132 (1980)
9. Rödder, W., Xu, L.: Elimination of Inconsistent Knowledge in the Probabilistic Expertsystem-Shell SPIRIT (in German). In: Operations Research Proceedings 2000. pp. 260–265. Springer-Verlag (2001)
10. Thimm, M.: Measuring Inconsistency in Probabilistic Knowledge Bases. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (2009)
11. Thimm, M.: Probabilistic Reasoning with Incomplete and Inconsistent Beliefs. Ph.D. thesis, Technische Universität Dortmund, Germany (2011), submitted

[2] http://sourceforge.net/projects/tweety/